DOUGLAS COLLEGE
COMMERCE AND BUSINESS ADMINISTRATION

Please implement **Regression models** on the ***Customer Churn*** (***Telecom Churn)*** dataset in ***R Studio***. You should try the best model to reach the best prediction. For detailed information about the implementation, please refer to the fifth lecture, on October 02, 2024.

The dataset is a binary-labeled dataset with two classes: 0 for customers who have not churned and 1 for those who have churned. You can download the dataset from the following link:

https://www.kaggle.com/datasets/barun2104/telecom-churn?resource=download

The deliverable for each group is a PDF file as the report with members' name. Your report should include the following tasks, along with screenshots of the codes and results and the explanation for each model respectively.

- ✓ **Partitioning**: Splitting the dataset into training and test sets.
- ✓ **Modeling**: Building different (at least four) regression models using the training set.
- ✓ **Regression Diagnostic** (Linearity, Normality, Homoscedasticity and Independence) and Outlier detection.
- ✓ **Interaction investigation based on correlation. Plotting The effects.**
- ✓ **Comparison**:
    - Comparing the models based on *ANOVA*
    - Comparing the models based on *AIC*
    - Comparing the models based on *Adjusted R square*
    - Comparing the models based on *CP*

	Excellent	Good	Needs Improvement
Partitioning (0.5 Points)	• Splitting the dataset into train and test sets • Paying attention to the partitioning type • Paying attention to the partitioning percentage. (0.5 Point)	• The partitioning type is not properly selected. • The partitioning percentage is not properly selected. (0.25 Point)	• No partitioning (the model is built based on the whole dataset) • Not enough or readable screenshots and explanations. (0 Point)
Building different regression models (at least 4 models) (2 Points)	✓ Making regression models. ✓ Analyzing the result of each model in order to make a better model. (2 Point)	Not enough analyzing and comparisons (1-1.75 Point)	Not enough or readable screenshots and explanations. (0 – 0.75 Point)
Interaction investigation based on correlation. Plotting The effects. (0.5 Points)	• Effects plot is clearly explained. • The interaction investigation is based on the correct selection of values in the correlation table. (0.5 Point)	The interaction is not based on the significant values in the result of correlation. (0.25 Point)	Not enough or readable screenshots and explanations. (0 Point)
Regression Diagnostic (1 Points)	• Linearity, • Normality, • Homoscedasticity and Independence • Outlier detection (1 Point)	No comparison among models. (0.75 Point)	Not enough or readable screenshots and explanations. (0 – 0.5 Point)
Comparison (1 Points)	• ANOVA • AIC • ADJUSTED R2 • CP (1 Point)	The plotting needs improvement. (0.75 Point)	Not enough or readable screenshots and explanations. (0 – 0.5 Point)